1C135 U.S. PTO 09/014422 09/014422 09/014422 01/27/98

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re A	Application of:)	
Masaki IWAMOTO, et al.)) Group Art Unit: To Be Assigne	
Serial	No.: To Be Assigned)	
Filed:	January 27, 1998	Examiner: To Be Assigned	
For:	INTERACTIVE DATA ANALYSIS SUPPO WHICH IS RECORDED AN INTERACT		

SUBMISSION OF CERTIFIED COPY OF PRIOR FOREIGN APPLICATION IN ACCORDANCE WITH THE REQUIREMENTS OF 37 C.F.R. § 1.55

Assistant Commissioner for Patents Washington, D.C. 20231

PROGRAM

Sir:

In accordance with the provisions of 37 C.F.R. § 1.55, Applicants submit herewith a certified copy of each of the following foreign applications:

Japanese Appln. No. 9-163692, filed June 20, 1997.

It is respectfully requested that Applicants be given the benefit of the earlier foreign filing date, as evidenced by the certified papers attached hereto, in accordance with the requirements of 35 U.S.C. § 119.

Respectfully submitted,

STAAS & HALSEY

Dated: January 27, 1998

By:

James D. Halsey, Jr. Registration No. 22,729

700 Eleventh Street, N.W. Suite 500 Washington, D.C. 20001 (202) 434-1500

日本国特許庁

PATENT OFFICE
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 Date of Application:

1997年 6月20日

出 願 番 号 Application Number:

平成 9年特許願第163692号

出 願 人 Applicant (s):

富士通株式会社

CERTIFIED COPY OF PRIORITY DOCUMENT



1997年11月21日

特許庁長官 Commissioner, Patent Office 荒井 寿 潭 順

特平 9 163692

【書類名】 特許願

【整理番号】 9704177

【提出日】 平成 9年 6月20日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/30

G06F 17/60

【発明の名称】 対話型データ分析支援装置及び対話型データ分析支援プ

ログラムを記録した媒体

【請求項の数】 18

【発明者】

【住所又は居所】 神奈川県小田原市栄町一丁目6番地4 株式会社富士通

ソフトウェア生産技術研究所内

【氏名】 岩本 昌己

【発明者】

【住所又は居所】 神奈川県小田原市栄町一丁目6番地4 株式会社富士通

ソフトウェア生産技術研究所内

【氏名】 本多 正人

【発明者】

【住所又は居所】 神奈川県小田原市栄町一丁目6番地4 株式会社富士通

ソフトウェア生産技術研究所内

【氏名】 伏見 俊彦

【発明者】

【住所又は居所】 神奈川県小田原市栄町一丁目6番地4 株式会社富士通

ソフトウェア生産技術研究所内

【氏名】 鈴木 輝幸

【発明者】

【住所又は居所】 神奈川県小田原市栄町一丁目6番地4 株式会社富士通

ソフトウェア生産技術研究所内

【氏名】 井上 征男

特平 9-163692

【発明者】

【住所又は居所】 神奈川県小田原市栄町一丁目6番地4 株式会社富士通

ソフトウェア生産技術研究所内

【氏名】

都築 康一

【発明者】

【住所又は居所】 神奈川県小田原市栄町一丁目6番地4 株式会社富士通

ソフトウェア生産技術研究所内

【氏名】

加藤 博己

【特許出願人】

【識別番号】

000005223

【氏名又は名称】

富士通株式会社

【代理人】

【識別番号】

100072590

【弁理士】

【氏名又は名称】

井桁 貞一

【電話番号】

044-754-3035

【手数料の表示】

【予納台帳番号】

011280

【納付金額】

21,000円

【提出物件の目録】

【物件名】

明細書 1

【物件名】

図面 1

【物件名】

要約書 1

【包括委任状番号】

9704486

【プルーフの要否】

要

【書類名】 明細書

【発明の名称】 対話型データ分析支援装置及び対話型データ分析支援プログラムを記録した媒体

【特許請求の範囲】

【請求項1】

データの分析を支援する対話型データ分析支援装置であって、

指定された集計条件に従って、分析対象のデータをクロス集計したクロス集計表を表示するクロス集計表表示手段と、

前記クロス集計表を構成する複数のセルから少なくとも1つのセルを指定する セル指定手段と、

前記セル指定手段により指定されたセルの範囲内で、分析対象のデータをグラフとして表示するグラフ表示手段と、

を含んで構成されることを特徴とする対話型データ分析支援装置。

【請求項2】

前記グラフ表示手段は、表示されるデータの範囲を限定する表示限定手段を備えたことを特徴とする請求項1記載の対話型データ分析支援装置。

【請求項3】

前記グラフ表示手段は、所定の条件に従って、表示されるデータを自動的に並べ替える並べ替え手段を備えたことを特徴とする請求項1又は2に記載の対話型 データ分析支援装置。

【請求項4】

前記グラフ表示手段は、前記分析対象のデータの特徴を抽出して新たな表示項目を見出す自動分析手段と、見出された新たな表示項目をグラフに追加する表示項目追加手段と、を含んで構成されることを特徴とする請求項1~3のいずれか1つに記載の対話型データ分析装置。

【請求項5】

前記グラフ表示手段は、該グラフ表示手段上で行われた操作の結果を、前記クロス集計表表示手段で使用される集計条件として保存する保存手段を備え、前記クロス集計表表示手段は、保存された集計条件に従って、分析対象のデータをク

ロス集計したクロス集計表を表示可能な構成であることを特徴とする請求項2~ 4のいずれか1つに記載の対話型データ分析装置。

【請求項6】

前記分析対象のデータは、複数のデータ項目から構成されるレコードの集合であることを特徴とする請求項1~5のいずれか1つに記載の対話型データ分析支援装置。

【請求項7】

前記グラフ表示手段は、前記分析対象のデータの中から所定数のレコードをランダムに抽出するランダム抽出手段を備え、抽出されたレコードに基づいてグラフを表示する構成であることを特徴とする請求項6記載の対話型データ分析支援 装置。

【請求項8】

前記グラフ表示手段は、前記データ項目を軸としてグラフを表示する構成であることを特徴とする請求項6又は7に記載の対話型データ分析支援装置。

【請求項9】

前記グラフ表示手段は、前記レコードを構成するデータ項目と同数の軸からなり、前記レコードの各々について、各データ項目の値に対応する点をプロットし、隣接する軸にプロットした点を線分で結んだグラフを表示する構成であることを特徴とする請求項8記載の対話型データ分析支援装置。

【請求項10】

データの分析を支援する対話型データ分析支援プログラムを記録した媒体であって、

指定された集計条件に従って、分析対象のデータをクロス集計したクロス集計表を表示するクロス集計表表示機能と、

前記クロス集計表を構成する複数のセルから少なくとも1つのセルを指定する セル指定機能と、

前記セル指定機能により指定されたセルの範囲内で、分析対象のデータをグラフとして表示するグラフ表示機能と、

を実現するプログラムを少なくとも記録したことを特徴とする対話型データ分

析支援プログラムを記憶した媒体。

【請求項11】

前記グラフ表示機能は、表示されるデータの範囲を限定する表示限定機能を備 えたことを特徴とする請求項10記載の対話型データ分析支援プログラムを記録 した媒体。

【請求項12】

前記グラフ表示機能は、所定の条件に従って、表示されるデータを自動的に並べ替える並べ替え機能を備えたことを特徴とする請求項10又は11に記載の対 話型データ分析支援プログラムを記録した媒体。

【請求項13】

前記グラフ表示機能は、前記分析対象のデータの特徴を抽出して新たな表示項目を見出す自動分析機能と、見出された新たな表示項目をグラフに追加する表示項目追加機能と、を含んで構成されることを特徴とする請求項10~12のいずれか1つに記載の対話型データ分析装置。

【請求項14】

前記グラフ表示機能は、前記グラフ上で行われた操作の結果を、前記クロス集計表表示機能で使用される集計条件として保存する保存機能を備え、前記クロス集計表表示機能は、保存された集計条件に従って、分析対象のデータをクロス集計したクロス集計表を表示可能であることを特徴とする請求項11~13のいずれか1つに記載の対話型データ分析プログラムを記録した媒体。

【請求項15】

前記分析対象のデータは、複数のデータ項目から構成されるレコードの集合で あることを特徴とする請求項10~14のいずれか1つに記載の対話型データ分 析支援プログラムを記録した媒体。

【請求項16】

前記グラフ表示機能は、前記分析対象のデータの中から所定数のレコードをランダムに抽出するランダム抽出機能を備え、抽出されたレコードに基づいてグラフを表示することを特徴とする請求項15記載の対話型データ分析支援プログラムを記録した媒体。

【請求項17】

前記グラフ表示機能は、前記データ項目を軸としてグラフを表示することを特徴とする請求項15又は16に記載の対話型データ分析支援プログラムを記録した媒体。

【請求項18】

前記グラフ表示機能は、前記レコードを構成するデータ項目と同数の軸からなり、前記レコードの各々について、各データ項目の値に対応する点をプロットし、隣接する軸にプロットした点を線分で結んだグラフを表示することを特徴とする請求項17記載の対話型データ分析支援プログラムを記録した媒体。

【発明の詳細な説明】

[0001]

【発明の属する技術分野】

本発明は、プログラムされたコンピュータによってデータの分析を支援する技術に関する。

[0002]

【従来の技術】

近年、データウェアハウスに格納された大量のデータを、より戦略的に活用しようというブームがある。このため、エンドユーザ向けツールとして、OLAP (オンライン分析処理)が認知され始めている。現状のOLAPは、クロス集計表を基本構成とし、ユーザ自身が、集計表の形でデータを整理しながらヒューリスティックに問題点を見つけ、問題点の原因を突き止めようとするものである。 具体的には、OLAPは、明確に定義された種々の切り口を自在に用い、各切り口毎に集計という手段で要約を行い、データを整理していくものである。

[0003]

ここで、切り口を自在に用いる手段として、ドリリング、スライシング及びダイシング等が用いられる。ドリリング、スライシング及びダイシング等によりデータを整理或いは分析する場合には、分析者は、問題点を認識した上で、クロス集計表のセルの値を理解し、問題点の原因となるものを洞察して、これを検証するための切り口で再度整理し直すというプロセスを繰り返す。

[0004]

【発明が解決しようとする課題】

しかしながら、クロス集計表のセルの値を理解し、問題点の原因となるものを 洞察するというプロセスは、分析者の経験やスキルに大きく影響される。また、 データの捉えかたや切り口の定義方法などは、分析者毎に異なる。さらに、分析 者がある発想を行った場合には、新しいデータを用意しなければならず、データ の分析を煩雑にする原因ともなっている。

[0005]

このため、これらの問題を解決する手段として、本願出願人は、切り口を示す データモデルをメタデータとして外付けで管理する方式、及び、個人毎に異なる 情報の切り口を柔軟に定義する"管理ポイント"という概念を適用する方式を提 案した(特開平8-180072号公報参照)。

しかし、このような工夫をしても解決できない問題点が依然としてある。

[0006]

即ち、集計という手段で要約する場合には、問題点を覆い隠してしまう傾向があり、問題点の発見すらできず、OLAPの利点を活かしきれない。また、クロス集計を単純化するしくみは、切り口を単純に増やし、問題点の原因を突き止めるための切り口の探索が極めて困難になってしまう。さらに、近年の市場やビジネスは常に変動するため、この変動に対応して切り口も変化させる必要がある。このため、データの分析にはより深い洞察が要求され、事前に用意された切り口による検証的なアプローチだけでは、かかる変動に対応することができない。

[0007]

そこで、このような問題点を発見的アプローチで解決しようとするデータマイニングが開発された。データマイニングは、データウェアハウスに格納された大量のデータを有効に活用する技術であり、有益なデータの自動抽出を行うものである。

しかし、データマイニングは、データウェアハウスに格納されたデータを有効 に活用する反面、次のような問題点がある。

[0008]

即ち、データマイニングの処理内容がブラックボックスであり、ユーザはどのような処理が行われているか全く不明である。また、データマイニングを使用するためには、数学的及びビジネス的な高度のスキルが要求され、エンドユーザが容易に使用できるものではない。さらに、問題点の原因らしきものが探索できたとしても、その検証が難しい。

[0009]

そこで、本発明は以上のような従来の問題点に鑑み、データウェアハウスに格納された大量のデータを、簡単な操作で自由に扱いうるGUI (Graphic User Interface) を提供することで、エンドユーザがデータの分析を効率的かつ容易に行えるようにすることを目的とする。

[0010]

【課題を解決するための手段】

このため、請求項1記載の発明は、データの分析を支援する対話型データ分析 支援装置を、指定された集計条件に従って、分析対象のデータをクロス集計した クロス集計表を表示するクロス集計表表示手段と、前記クロス集計表を構成する 複数のセルから少なくとも1つのセルを指定するセル指定手段と、前記セル指定 手段により指定されたセルの範囲内で、分析対象のデータをグラフとして表示す るグラフ表示手段と、を含んで構成した。

[0011]

ここで、「クロス集計」とは、少なくとも1行及び1列からなる2次元配列において、特定行或いは特定列に対して合計、標準偏差等の数学的、統計的な処理を行うことをいう。また、「セル」とは、クロス集計表の特定行及び特定列によって定義される最小単位の構成要素をいう。

かかる構成によれば、分析対象のデータをクロス集計したクロス集計表を構成 する複数のセルから少なくとも1つのセルを指定した後、指定したセルの範囲内 で、必要に応じて分析対象のデータがグラフとして表示される。従って、問題点 の発生原因の探索を進めようとする条件を満たすデータだけがグラフとして表示 されるので、分析者は、データの分析を視覚的に行うことが可能となる。

[0012]

請求項2記載の発明は、前記グラフ表示手段は、表示されるデータの範囲を限 定する表示限定手段を備えた構成とした。

かかる構成によれば、表示限定手段によって、表示されるデータの範囲が限定 されるので、問題点の原因であると思える範囲内のデータのみを表示することが でき、問題点の原因の探索においてその範囲を徐々に限定していくことで、最終 的な原因を容易に見出すことが可能となる。

[0013]

請求項3記載の発明は、前記グラフ表示手段は、所定の条件に従って、表示されるデータを自動的に並べ替える並べ替え手段を備えた構成とした。

かかる構成によれば、並べ替え手段によって、グラフに表示されるデータが色々な順序で並べ替えられるので、データの表示順序により問題点の原因が明確になる場合がある。

[0014]

請求項4記載の発明は、前記グラフ表示手段は、前記分析対象のデータの特徴を抽出して新たな表示項目を見出す自動分析手段と、見出された新たな表示項目をグラフに追加する表示項目追加手段と、を含んで構成した。

かかる構成によれば、自動分析手段によって見出された表示項目を、表示項目 追加手段によってグラフに追加することで、問題点の原因が明確になる場合があ る。

[0015]

請求項5記載の発明は、前記グラフ表示手段は、該グラフ表示手段上で行われた操作の結果を、前記クロス集計表表示手段で使用される集計条件として保存する保存手段を備え、前記クロス集計表表示手段は、保存された集計条件に従って、分析対象のデータをクロス集計したクロス集計表を表示可能な構成とした。

かかる構成によれば、保存手段によって、グラフ表示手段上で行われた操作の結果が集計条件として保存され、この集計条件を利用して分析対象のデータをクロス集計したクロス集計表が表示可能となる。従って、問題点の原因探索において行った操作が、次回の原因探索でも利用することができるようになり、分析処理効率を飛躍的に向上することが可能となる。

[0016]

請求項6記載の発明は、前記分析対象のデータは、複数のデータ項目から構成 されるレコードの集合である構成とした。

かかる構成によれば、分析対象のデータは、同一のデータ項目から構成される レコードの集合として取り扱うことができるので、関連性のあるレコードを簡単 な操作で一括して取り扱うことが可能となる。

[0017]

請求項7記載の発明は、前記グラフ表示手段は、前記分析対象のデータの中から所定数のレコードをランダムに抽出するランダム抽出手段を備え、抽出された レコードに基づいてグラフを表示する構成とした。

かかる構成によれば、ランダム抽出手段によって所定数のレコードのみがランダムに抽出され、抽出されたレコードに基づいてグラフが表示されるので、グラフの表示のために使用されるレコード数が減少する。従って、データベース等からレコードを読み取る時間が短縮し、グラフ表示手段において高速に表示を行うことが可能となる。

[0018]

請求項8記載の発明は、前記グラフ表示手段は、前記データ項目を軸としてグラフを表示する構成とした。

かかる構成によれば、データ項目の値が軸上に表わされるので、その値の大小 を直感的に把握することができ、データの傾向の把握を容易に行うことが可能と なる。

[0019]

請求項9記載の発明は、前記グラフ表示手段は、前記レコードを構成するデータ項目と同数の軸からなり、前記レコードの各々について、各データ項目の値に対応する点をプロットし、隣接する軸にプロットした点を線分で結んだグラフを表示する構成とした。

かかる構成によれば、各レコードのデータ項目間の相関関係、及び、レコード間の相関関係が視覚的に表示されるので、問題点の原因の探索を容易に行うことが可能となる。

[0020]

請求項10記載の発明は、データの分析を支援する対話型データ分析支援プログラムを記録した媒体を、指定された集計条件に従って、分析対象のデータをクロス集計したクロス集計表を表示するクロス集計表表示機能と、前記クロス集計表を構成する複数のセルから少なくとも1つのセルを指定するセル指定機能と、前記セル指定機能により指定されたセルの範囲内で、分析対象のデータをグラフとして表示するグラフ表示機能と、を含んで構成した。

[0021]

ここで、「媒体」とは、各種情報を確実に記録でき、かつ、必要に応じて確実に取り出すことが可能なものをいい、具体的には、紙カード(パンチカード),紙テープ、磁気テープ、磁気ディスク、磁気ドラム、ICカード、CD-ROM等が該当する。

かかる構成によれば、請求項1記載の発明と同様な作用に加えて、分析対象のデータをクロス集計したクロス集計表を表示するクロス集計表表示機能、クロス集計表を構成する複数のセルから少なくとも1つのセルを指定するセル指定機能、及び、指定されたセルの範囲内で分析対象のデータをグラフとして表示するグラフ表示機能を実現するためのプログラムを媒体に記録したので、これらのプログラムを記録した媒体があれば、多数の計算機に対して前記各機能を持たせることが可能となる。

[0022]

請求項11記載の発明は、前記グラフ表示機能は、表示されるデータの範囲を 限定する表示限定機能を備えるようにした。

かかる構成によれば、表示限定機能によって、表示されるデータの範囲が限定 されるので、問題点の原因であると思える範囲内のデータのみを表示することが でき、問題点の原因の探索においてその範囲を徐々に限定していくことで、最終 的な原因を容易に見出すことが可能となる。

[0023]

請求項12記載の発明は、前記グラフ表示機能は、所定の条件に従って、表示 されるデータを自動的に並べ替える並べ替え機能を備えるようにした。 かかる構成によれば、並べ替え機能によって、グラフに表示されるデータが色々な順序で並べ替えられるので、データの表示順序により問題点の原因が明確になる場合がある。

[0024]

請求項13記載の発明は、前記グラフ表示機能は、前記分析対象のデータの特徴を抽出して新たな表示項目を見出す自動分析機能と、見出された新たな表示項目をグラフに追加する表示項目追加機能と、を含んで構成した。

かかる構成によれば、自動分析機能によって見出された表示項目を、表示項目 追加機能によってグラフに追加することで、問題点の原因が明確になる場合があ る。

[0025]

請求項14記載の発明は、前記グラフ表示機能は、前記グラフ上で行われた操作の結果を、前記クロス集計表表示機能で使用される集計条件として保存する保存機能を備え、前記クロス集計表表示機能は、保存された集計条件に従って、分析対象のデータをクロス集計したクロス集計表を表示可能である構成とした。

かかる構成によれば、保存機能によって、グラフ表示手段上で行われた操作の結果が集計条件として保存され、この集計条件を利用して分析対象のデータをクロス集計したクロス集計表が表示可能となる。従って、問題点の原因探索において行った操作が、次回の原因探索でも利用することができるようになり、分析処理効率を飛躍的に向上することが可能となる。

[0026].

請求項15記載の発明は、前記分析対象のデータは、複数のデータ項目から構成されるレコードの集合とした。

かかる構成によれば、分析対象のデータは、同一のデータ項目から構成される レコードの集合として取り扱うことができるので、関連性のあるレコードを簡単 な操作で一括して取り扱うことが可能となる。

[0027]

請求項16記載の発明は、前記グラフ表示機能は、前記分析対象のデータの中から所定数のレコードをランダムに抽出するランダム抽出機能を備え、抽出され

たレコードに基づいてグラフを表示するようにした。

かかる構成によれば、ランダム抽出機能によって、所定数のレコードのみがランダムに抽出され、抽出されたレコードに基づいてグラフが表示されるので、グラフの表示のために使用されるレコード数が減少する。従って、グラフを表示する際に、データベース等からレコードを読み取る時間が短縮し、グラフ表示機能において高速に表示を行うことが可能となる。

[0028]

請求項17記載の発明は、前記グラフ表示機能は、前記データ項目を軸として グラフを表示するようにした。

かかる構成によれば、データ項目の値が軸上に表わされるので、その値の大小 を直感的に把握することができ、データの傾向の把握を容易に行うことが可能と なる。

[0029]

請求項18記載の発明は、前記グラフ表示機能は、前記レコードを構成するデータ項目と同数の軸からなり、前記レコードの各々について、各データ項目の値に対応する点をプロットし、隣接する軸にプロットした点を線分で結んだグラフを表示するようにした。

かかる構成によれば、各レコードのデータ項目間の相関関係、及び、レコード 間の相関関係が視覚的に表示されるので、問題点の原因の探索を容易に行うこと が可能となる。

[0030]

【発明の実施の形態】

以下、添付された図面を参照して本発明を詳述する。

先ず、本発明の概要を説明すると、情報モデルがデータと独立したメタデータとして定義された方式で動作するOLAP等のクロス集計表作成機能をベースとして、集計されたデータの意味と原因探索の方向付けを、平行座標をベースにしたビジュアライザ(詳細は後述する)等のグラフ表示機能によって行えるようにする。即ち、クロス集計表作成機能とグラフ表示機能とが連係動作するようにし、分析者たるユーザが、クロス集計表作成機能或いはグラフ表示機能の一方を操

作すると、この操作が他方に反映されるようにすることで、直感的にデータの分析作業が行えるようにする。なお、以下で説明する実施形態では、データウェアハウスに格納された大量のデータを有効に活用するため、データマイニング手法を利用している。

[0031]

図1及び図2は、本発明に係る対話型データ分析支援装置を、汎用的な計算機 システムの上で実現した一実施形態のシステム構成を示す。

ここで、汎用的な計算機システムとしては、パーソナルコンピュータやワークステーション等の計算機が該当する。即ち、計算機システム10は、図2に示すように、命令の解釈と実行とを制御する中央処理装置(以下「CPU」という)11と、CPU11で処理された処理結果データやプログラム等を記憶している主記憶装置12と、計算機システム10の外部からデータを取り込み、主記憶装置12にデータを送り込むキーボード等の入力装置13と、主記憶装置12に記憶されている処理結果データを、計算機システム10外に取り出すCRT等の出力装置14と、プログラムやデータを長期間に亘って保存する磁気ディスク等の補助記憶装置15と、主記憶装置12と周辺機器(入力装置13、出力装置14及び補助記憶装置15)との間でデータの授受を行う入出力チャネル16と、を含んで構成される。

[0032]

そして、CD-ROM等の可搬型の記憶媒体に格納された対話型データ分析支援プログラムを、例えば、CD-ROMドライブから補助記憶装置15としての磁気ディスクにインストールした上で、対話型データ分析支援プログラムが実行される。

図1は、データウェアハウスをサーバに、OLAPをクライアントにしたクライアント/サーバ分散方式のシステム構成を示す。即ち、サーバは、ディスク上のデータを管理し、クライアントからの処理要求を処理する。このようにすれば、異なるアーキテクチャ上でクライアントが生成したデータウェアハウスにもアクセスが可能となる。

[0033]

クライアント20は、OLAPクライアント部21と、ビジュアライザ部22 と、データマイニング部23と、概観情報要約部24と、を含んで構成される。

OLAPクライアント部21は、図3に示すように、ユーザが設定した管理ポイントとしての表側項目及び表頭項目からなるクロス集計表を作成するOLAP 21aから構成される。OLAP21aは、所定の表側項目及び表頭項目に対応するセルに数値を入力して、その集計を行うものである。なお、管理ポイントとしては、表側項目及び表頭項目だけではなく、他の項目も設定できるようになっている。

[0034]

ここで、OLAP21aにおいてデータを整理していく際に、「切り口の定義」,「軸の定義」及び「切り口の階層定義」という手法が利用される。即ち、「切り口の定義」とは、文字データやカテゴリデータを任意のくくりでグループ化したり、或いは、数値データを任意の範囲で括ることをいう。例えば、東京都、神奈川県、千葉県等を"関東地区"という切り口にすることが該当する。「軸の定義」とは、所定のデータ項目で任意の切り口の定義を行った結果、意味的に新しい軸を定義することをいう。例えば、"年齢"という軸を、10歳毎に区切って"世代"という新しい軸を定義することが該当する。「切り口の階層定義」とは、所定の切り口で定義した切り口を基に、さらに新しい切り口を定義して上位概念とすることをいう。例えば、渋谷、新宿、東京を"東京"とすると共に、横浜、川崎を"神奈川"という切り口で定義した上で、"東京", "神奈川"を"関東地区"という切り口で定義することが該当する。そして、OLAP21aでは、かかる切り口を自在に用い、各切り口毎に集計という手段で要約を行い、データを整理して問題点の原因を突き止めていく。

[0035]

ビジュアライザ部22は、図4に示すように、OLAPクライアント部21で作成されたクロス集計表のデータを、平行座標からなるグラフ上にプロットして、レコード毎の線分を重ねて表示する機能を有している。具体的には、平行座標からなるグラフを表示すると共に、ユーザに対してフィルタリング(詳細は後述する)等の操作を行わせるビジュアライザ22aと、ビジュアライザ22aで使

用するデータを保持するビジュアライザ用データバッファ22bと、を含んで構成される。また、ビジュアライザ用データバッファ22bは、OLAPクライアント部21で定義された管理ポイントの個数に応じて、少なくとも1つの2次元配列から構成される。

[0036]

ビジュアライザ22aは、平行座標からなるグラフを表示する際に、ユーザがマクロな傾向や特徴、及び、問題点の原因が類推できるように、少なくとも、以下の情報を表示する機能を有している。

- (1) クロス集計表の管理ポイントとしての表側項目及び表頭項目
- (2) 他の管理ポイント
- (3) ランダムサンプリングされた全項目の詳細値
- (4) 全管理ポイントの組み合わせにおける、件数或いは合計等の概集計値
- (5) 異常データ(特異なデータを含む)

ここで、全管理ポイントの組み合わせにおける集計値を含む理由としては、例えば、クロス集計表のあるセルの値が大きくなっている原因を調べる場合、クロス集計表上の集計値が異常値となっている原因として、次の原因が考えられる。即ち、異常な値の詳細データが混入していて、その値に起因してあるセルの値が大きくなっているか、或いは、対象データ件数が大きいため集計値が大きくなっている。このような状態をユーザが容易に認識できるように、異常詳細データと対象データ件数等の統計量を同時に表示している。

[0037]

この場合、統計量は概略の統計量でよい。即ち、後述するデータウェアハウス32に格納される大量のデータの全レコードにアクセスする必要はなく、ランダムサンプリングされたビジュアライザ用データバッファ22bに含まれるデータの統計量で足りる。なぜならば、ユーザは、表示されるデータの相対的な傾向或いは特徴を把握できれば、データ分析を充分遂行できるからである。また、どういう条件のデータを異常値として扱うかも、ランダムサンプリングされたデータに基づいて判断し、判断された条件の元でデータウェアハウス32から抽出するようにする。

[0038]

ビジュアライザ22a上では、表示限定手段又は機能として、マウス等のポインティングデバイスによるドラッグ操作によって、各軸毎にフィルタリングをすることができる。即ち、軸の上端部或いは下端部をクリックして、表示したいと所望する範囲までドラッグすると、指定した範囲内に含まれるデータを表わす線分のみが、色や輝度等により強調表示される。従って、ユーザは強調表示された線分を検討することで、問題点の原因探索が直感的に行えることとなる。

[0039]

また、ビジュアライザ22aには以下の機能が要求される。

- (1) クロス集計表自体を条件選択ビュー、即ち、クロス集計表の複数セルから 少なくとも1つのセルを指定して(セル指定手段又は機能)、選択された条件下 で詳細データをビジュアルに表示できること。
 - (2) 高速に表示及びフィルタリング操作が行えること。

[0040]

- (3) 情報全体の傾向(トレンド)と異常値とを容易に把握できること。
- (4) 傾向を把握する際に、意味のない情報は自動的に非表示となり、ノイズのない情報が提示されること。また、階層的に関連性を持った情報は、ドリリング操作により階層的に詳細化できること。

ここで、重要度の低いデータを非表示にする機構について説明すると、管理ポイントの中で階層的に定義されている項目については、第1階層のみを初期表示し、ユーザからの指示によって、下位の階層を順次表示する。上位階層で範囲が限定されている場合には、下位の階層では、選択された上位階層の範囲内のデータだけを表示する(ドリリング)。また、データ項目間の相関をとり、非常に高い相関を持つものを非表示にする。或いは、主成分を新しい軸として表示し、下位階層に元の項目を展開する。さらに、名義尺度のデータについては、各データの出現頻度やエントロビー(情報量)によって、価値を判断する。

[0041]

(5) 線分の類似性から自動的に新たな分類ができること(自動分析手段又は機能)。そして、分類した結果として新しい分類軸が追加され(表示項目追加手段

又は機能)、簡単なフィルタリング操作により各分類の選択表示ができること。

即ち、データマイニング手法を活用することによって、分析対象のデータの特徴を自動抽出した後、新たな管理ポイントである分類軸を見出し、これをビジュアライザ22a上に追加できること。

[0042]

- (6) 特定のグループに関連性の強い各軸の項目を、所定の条件に従って、自動 的に近くに再配置して表示ができること(並べ替え手段又は機能)。
- (7) ビジュアライザ22a上で整理した情報を、メタデータとして即座に格納できること(保存手段又は機能)。そして、その情報は、OLAP21aにより切り口として利用できること。

[0043]

(8) ビジュアライザ22a上でフィルタリングした結果をOLAP21aに通知して、ビジュアライザ22a上の検証だけではなく、OLAP21aと連動した精度の高い検証作業ができること。

データマイニング部23は、ビジュアライザ用データバッファ22bから有益なデータのみを自動抽出する機能を有している。具体的には、データマイニング処理の各種制御を行うマイニング制御部23aと、実際にデータマイニング処理を行うマイニングエンジン23bと、データマイニング処理を行うときに使用するデータを一時的に保持するマイニング用データバッファ23cと、マイニング用データバッファ23cと、マイニングータ変換を行うデータ変換部23dと、を含んで構成される。

[0044]

概観情報要約部24は、主に、ビジュアライザ部22でグラフを表示する際に、サーバ30側のデータウェアハウス32からデータをランダムサンプリングすることで、検索件数の減少による処理速度の向上を図るものである。即ち、ランダム抽出手段又は機能として作用する。このようにランダムサンプリングした結果は、統計的にマクロな傾向や特徴を保持しているため、ユーザは、ビジュアライザ部22で表示されたデータを、ユーザが認識している全情報を表わしていると考えても何ら問題がない。

[0045]

また、概観情報要約部24は、ランダムサンプリングされたデータを、全ての管理ポイント毎に集計する機能を有している。さらに、概観情報要約部24は、ランダムサンプリングされたデータから、各軸毎にヒンジとフェンスの値を計算して外挿値境界の条件を算出する機能を有している。そして、外挿値境界の条件を後述する概観情報収集部34に通知して、異常詳細データのみを抽出することもできる。

[0046]

一方、サーバ30は、OLAPサーバ部31と、データウェアハウス32と、 データマイニング部33と、概観情報収集部34と、を含んで構成される。

OLAPサーバ部31は、クライアント20側のOLAPクライアント部21 からの処理要求に対応する機能を有している。具体的には、各種処理の制御を行うサーバソフト31aと、データの構成、内容、キー及びインデックス等を定義したメタデータが格納されるメタデータファイル31bと、データウェアハウス32からデータを抽出する際に作業用データベースとなる集計結果ファイル31 cと、メタデータファイル31bに格納されたメタデータを加味しながら集計処理を行う集計エンジン31dと、を含んで構成される。

[0047]

データウェアハウス32は、意思決定支援システム機能をサポートするために 設計されたもので、統合化されたサブジェクト(主題)を指向したデータベース の集合である。そして、データベースには、複数のデータ項目から構成されるレ コードが複数集められ、1つの分析対象データが構成されている。

データマイニング部33は、データウェアハウス32から有益なデータのみを自動抽出する機能を有している。具体的には、データマイニング処理の制御を行うマイニング制御部33aと、実際にデータマイニング処理を行うマイニングエンジン33bと、データマイニング処理を行うときに使用するデータを一時的に保持するマイニング用データバッファ33cと、マイニング用データバッファ33cと、マイニング用データバッファ33cとデータウェアハウス32等との間のデータ変換を行うデータ変換部33dと、を含んで構成される。

[0048]

概観情報収集部34は、データウェアハウス32からデータをランダムサンプリングして、クライアント20側の概観情報要約部24に引き渡す機能を有している。

なお、OLAPクライアント部21及びOLAPサーバ部31を含んでクロス 集計表表示手段又は機能が実現され、また、クライアント20側のビジュアライ ザ部22, データマイニング部23, 概観情報要約部24、及び、サーバ30側 のデータマイニング部33, 概観情報収集部34を含んでグラフ表示手段又は機 能が実現される。

[0049]

次に、対話型データ分析支援装置の処理内容について、図5~図14の処理流 れ図及びフローチャートを参照しつつ説明する。

図5~図7は、OLAP21a上にクロス集計の集計結果を表示するまでの処理流れ、及び、処理内容を説明するフローチャートを示している。即ち、OLAP21aを操作中のユーザが、切り口を指定してクロス集計表の作成指示を行ったときの処理内容を示すものである。

[0050]

図6は、OLAP21a上に集計結果を表示するまでのクライアント20側における処理内容の詳細を示したフローチャートである。このルーチンは、ユーザがクロス集計表作成指示を行ったときに実行される。

ステップ1 (図では「S1」と略記する。以下同様)では、OLAPクライアント部21が、サーバ30側のサーバソフト31aにクロス集計表作成指示を送付する(処理A)。

[0051]

ステップ2では、サーバ30側のOLAPサーバ部31から集計結果が返送されたか否かが判断され、集計結果が返送されるまで待機する。

ステップ3では、返送された集計結果がOLAP21a上に表示される。

図7は、OLAP21a上に集計結果を表示するまでのサーバ30側における 処理内容の詳細を示したフローチャートである。このルーチンは、クライアント 20側からクロス集計表作成指示を受け取ったときに実行される。

[0052]

ステップ10では、クロス集計表作成指示を受け取ったサーバソフト31aが 、集計エンジン31dを起動する(処理B)。

ステップ11では、集計エンジン31 dが、データウェアハウス32から対象となるデータを読み込む(処理C)。

ステップ12では、集計エンジン31dが、メタデータファイル31bからメタデータを読み込み、メタデータを加味しながら集計処理を行う(処理D)。

[0053]

ステップ13では、集計エンジン31 dが、メタデータを加味した集計結果を 作業用の集計結果ファイル31 c に格納する(処理E)。

ステップ14では、集計結果ファイル31cに格納された集計結果が、クライアント20側のOLAPクライアント部21に返送される(処理F)。

図8及び図9は、OLAP21aのクロス集計表上に集計結果が表示されている状態において、ユーザが指定した範囲内のクロス集計表のデータをグラフ表示するまでの処理流れ、及び、処理内容を説明するフローチャートを示している。即ち、ユーザがクロス集計表の一部のセルの結果の意味付けを知るために、その部分をマウス等のポインティングデバイスで指定して、グラフ表示を指定したときに実行されるものである。

[0054]

ステップ20では、検索情報として、クロス集計表を作成するために使用した データベース名、管理ポイント及び検索条件が、OLAPクライアント部21か ら概観情報要約部24に送付される(処理A)。

ステップ21では、概観情報要約部24が、送付された検索情報に基づいて、サーバ30側の概観情報収集部34を介して、データウェアハウス32から詳細情報,要約情報及び異常詳細情報をランダムサンプリングして取得する(処理B)。このとき、ランダムサンプリングを行う個数を、ユーザが設定できるようにしてもよい。

[0055]

ステップ22では、概観情報要約部24が、メタデータファイル31bに格納 されているメタデータを取得して付加情報を作成し、取得した詳細情報、要約情 報及び異常詳細情報に付加情報を付加する(処理C)。

ステップ23では、概観情報要約部24が、付加情報が付加された詳細情報, 要約情報及び異常詳細情報に基づき、ビジュアライザ用データバッファ22bを 作成する(処理D)。

[0056]

ステップ24では、ビジュアライザ22aを起動し、ビジュアライザ用データ バッファ22bに展開されているデータを、平行座標からなるグラフにレコード 毎の線分として表示する(処理E)。

図10~図14は、ユーザが、ビジュアライザ22a上で表示内容を更新する 処理、例えば、データの自動分類や軸項目の並べ換えの指示を行った場合に、実 行される処理流れ、及び、処理内容を説明するフローチャートを示している。表 示内容を更新する処理は、対象データ量が所定量より少ない場合にはクライアン ト20側で行われ、対象データ量が所定量以上の場合にはサーバ30側で行われ る。

[0057]

図10及び図11は、クライアント20側でビジュアライザ22aの表示内容の更新処理が行われるときの処理内容を示す。

ステップ30では、表示内容を更新する指示内容が、データマイニング部23 のマイニング制御部23aに送付される(処理A)。

ステップ31では、マイニング制御部23aが、クライアント20側のマイニングエンジン23bを起動する(処理B)。

[0058]

ステップ32では、マイニングエンジン23bが、データ変換部23dを介して、ビジュアライザ用データバッファ22bのデータを変換し、マイニングエンジン入力用のマイニング用データバッファ23cを作成する(処理C)。

ステップ33では、マイニングエンジン23bが、マイニング用データバッファ23cからデータを入力し、データマイニング処理を行う(処理D)。

[0059]

ステップ34では、マイニング制御部23aが、マイニングエンジン23bにより行われたデータマイニング処理の結果を、データ変換部23dを介して、ビジュアライザ用データに変換し、変換した処理結果に基づいてビジュアライザ用データバッファ22bの内容を更新する(処理E)。

ステップ35では、ビジュアライザ22aが、ビジュアライザ用データバッファ22bのデータに基づいて表示内容を更新する(処理F)。この結果、データの自動分類或いは軸項目の並べ換え結果が、ビジュアライザ22a上に表示されることとなる。

[0060]

ステップ36は、ユーザが、ビジュアライザ22a上の結果を新しい情報の切り口として、メタデータとして保管するか否かの選択処理が行われる。そして、保管する場合にはステップ37へと進み、保管しない場合には本ルーチンを終了する。

ステップ37では、概観情報要約部24が、ビジュアライザ用データバッファ 22bのデータを、メタデータ形式に変換し、サーバ30側のサーバソフト31 aを介して、メタデータファイル31bにこれを格納する(処理G)。

[0061]

図12~図14は、サーバ30側でビジュアライザ22aの表示内容の更新処理が行われるときの処理内容を示す。具体的には、図12は、全体の処理流れ図、図13は、クライアント20側の処理内容を説明するフローチャート、図14は、サーバ30側の処理内容を示すフローチャートである。

先ず、クライアント20側の処理内容について説明する。

[0062]

ステップ40では、表示内容を更新する指示内容が、データマイニング部23 のマイニング制御部23aに送付される(処理A)。

ステップ41では、マイニング制御部23aが、検索情報としてのクロス集計表を作成するために使用したデータベース名、管理ポイント及び検索条件を、サーバ30側のマイニング制御部33aに送付する(処理B)。

[0063]

ステップ42では、サーバ30側のマイニング制御部33aから、データマイニング処理の結果が格納されたデータベース名及び管理ポイントが返送されたか否かが判断され、返送されるまで待機する。

ステップ43では、マイニング制御部23aが、返送されたデータベース名及び管理ポイントを、概観情報要約部24に送付する(処理H)。

[0064]

ステップ44では、概観情報要約部24が、サーバ30側のメタデータファイル31bに格納されているメタデータを取得して付加情報を作成し、返送されたデータベース名及び管理ポイントに基づき、データウェアハウス32から取得した詳細情報、要約情報及び異常詳細情報に付加情報を付加する(処理I)。

ステップ45では、概観情報要約部24が、付加情報が付加された詳細情報, 要約情報及び異常詳細情報に基づいて、ビジュアライザ用データバッファ22b の内容を更新する(処理 J)。

[0065]

ステップ46では、ビジュアライザ22aが、ビジュアライザ用データバッファ22bのデータに基づいて表示内容を更新する(処理K)。この結果、データの自動分類或いは軸項目の並べ換え結果が、ビジュアライザ22a上に表示されることとなる。

次に、サーバ30側の処理内容について説明する。

[0066]

ステップ50では、マイニング制御部33aが、クライアント20側から送付された検索情報としてのデータベース名、管理ポイント及び検索条件に基づいて、データ変換部33dを介して、データウェアハウス32から対象データを読み込み、マイニング用データバッファ33cを作成する(処理C)。

ステップ51では、マイニング制御部33aが、マイニングエンジン33bを 起動する(処理D)。

[0067]

ステップ52では、マイニングエンジン33bが、マイニング用データバッフ

ァ33cのデータを入力し、データマイニング処理を行う(処理E)。

ステップ53では、マイニング制御部33aが、マイニングエンジン33bで行われたデータマイニング処理の結果を、データ変換部33dを介してデータベース形式に変換し、変換されたデータマイニング処理の結果を、データウェアハウス32に格納する(処理F)。

[0068]

ステップ54では、マイニング制御部33aが、データマイニング処理の結果が格納されたデータベース名及び管理ポイントを、クライアント20側のマイニング制御部23aに返送する(処理G)。

このようにすれば、エンドユーザ向けのOLAPにおいて、問題点の原因を探索しようとする範囲を指定した後で、かかる範囲内における分析対象データを、平行座標からなるグラフ上に表示することができるようになる。そして、表示されたグラフ上において、フィルタリングや自動分類等の操作を行うことで、問題点の原因を次第に明らかにすることができる。従って、高度な経験やスキルを持っていなくとも、新しい発想や高い確度をもったクロス集計表の分析が可能となる。

[0069]

また、グラフ表示を行う際に、分析対象となるデータの全レコードを取り扱わず、所定数のレコードをランダムにサンプリングして使用するため、例えば、データベースへのアクセス件数の減少により処理時間が短縮し、グラフの表示が高速で行われるようになる。このように、レコードをランダムサンプリングして取り扱うデータ量を減らしても、グラフは、マクロなデータの傾向及び特徴を表わしているので、データの分析処理に対して何ら影響を及ぼさず、高速表示による思考の中断が防止され、極めて効果的な分析作業が可能となる。

[0070]

要するに、エンドユーザがデータマイニングに関して高度な知識を持たなくと も、もともとその分野で深い洞察力を有する業務担当者が、問題点の原因探索に おいてデータマイニング手法を全く意識せずに利用することができるようになる

[0071]

【発明の効果】

以上説明したように、請求項1又は請求項10に記載の発明によれば、分析対象のデータをクロス集計したクロス集計表において、問題点の発生原因の探索を進めようとする範囲を指定してグラフ上に表示することができるので、データの分析が視覚的に行なわれ、データの分析を効率的かつ容易に行うことができる。

[0072]

特に、請求項10記載の発明によれば、請求項1記載の発明の効果に加え、媒体にはクロス集計表表示機能、セル指定機能及びグラフ表示機能を実現するプログラムが記録されているので、本発明に係る対話型データ分析支援プログラムを流通させることができ、かかる媒体を取得した者は対話型データ分析支援装置を容易に構築することができる。

[0073]

請求項2又は請求項11に記載の発明によれば、問題点の原因の探索過程において、分析対象のデータの範囲を徐々に限定していくことで、最終的な原因を容易に見出すことができ、データ分析効率を向上することができる。

請求項3又は請求項12に記載の発明によれば、データの表示順序により問題 点の原因が明確になる場合があるので、データ分析効率を向上することができる

[0074]

請求項4又は請求項13に記載の発明によれば、自動分析により見出された表示項目がグラフに追加されることで、問題点の原因が明確になる場合があるので、データ分析効率を向上することができる。

請求項5又は請求項14に記載の発明によれば、問題点の原因探索において行った操作が、次回の原因探索でも利用することができるようになり、分析処理効率を飛躍的に向上することができる。

[0075]

請求項6又は請求項15に記載の発明によれば、関連性のあるレコードを簡単な操作で一括して取り扱うことが可能となり、操作性を向上することができる。

請求項7又は請求項16に記載の発明によれば、使用レコードの減少により処理時間が短縮し、グラフの表示が高速に行われるようになるので、システムに高速な処理能力がなくとも、データ分析処理にあたり充分な動作速度が確保でき、コスト上昇を抑制することができる。

[0076]

請求項8又は請求項17に記載の発明によれば、データの傾向の把握を容易に 行うことが可能となり、データ分析において、分析者の誤認等の発生が低減し、 分析結果の精度を向上することができる。

請求項9又は請求項18に記載の発明によれば、問題点の原因の探索を容易に 行うことが可能となり、データ分析効率を向上することができる。

【図面の簡単な説明】

【図1】

本発明の一実施形態を示すシステム構成図である。

【図2】

同上の計算機システムの詳細構成図である。

【図3】

OLAPに表示されるクロス集計表の一例を示す図である。

【図4】

ビジュアライザに表示されるグラフの一例を示す図である。

【図5】

クロス集計表の表示処理を示す処理流れ図である。

【図6】

クライアント側におけるクロス集計表の表示処理を示すフローチャートである

【図7】

サーバ側におけるクロス集計表の表示処理を示すフローチャートである。

【図8】

グラフの表示処理を示す処理流れ図である。

【図9】

グラフの表示処理を示すフローチャートである。

【図10】

クライアント側でグラフの表示更新処理を行う場合の処理流れ図である。

【図11】

クライアント側でグラフの表示更新処理を行う場合のフローチャートである。

【図12】

サーバ側でグラフの表示更新処理を行う場合の処理流れ図である。

【図13】

クライアント側におけるグラフの表示更新処理を示すフローチャートである。

【図14】

サーバ側におけるグラフの表示更新処理を示すフローチャートである。

【符号の説明】

21···OLAPクライアント部

22・・・ビジュアライザ部

23・・・データマイニング部

24・・・概観情報要約部

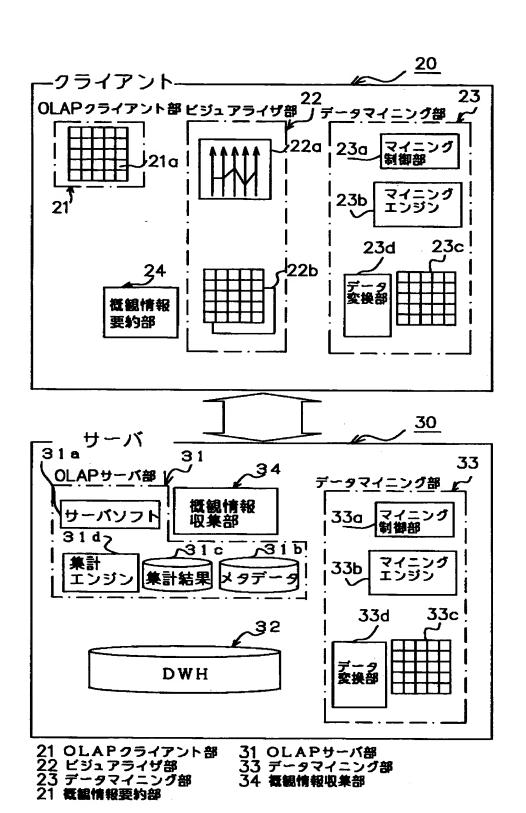
31···OLAPサーバ部

33・・・データマイニング部

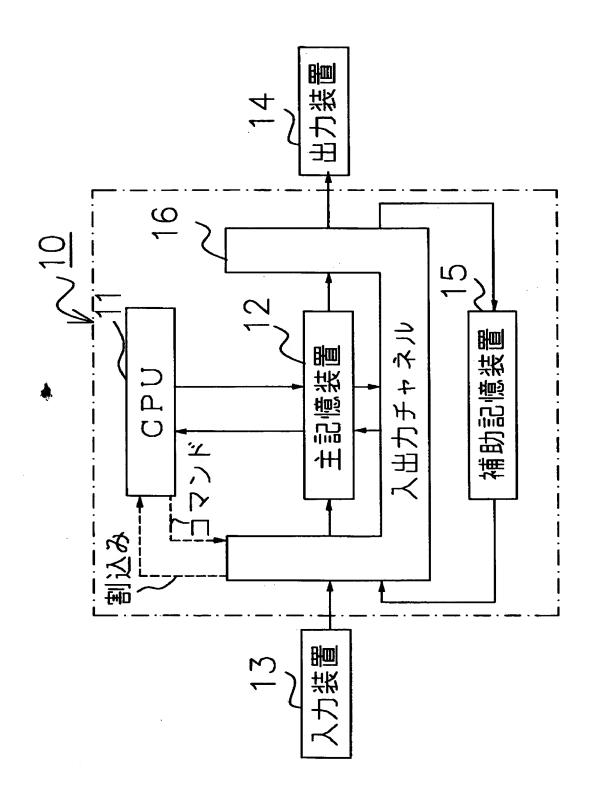
34・・・概観情報収集部

【書類名】 図面

【図1】



【図2】

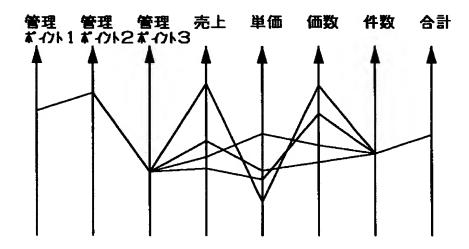


【図3】

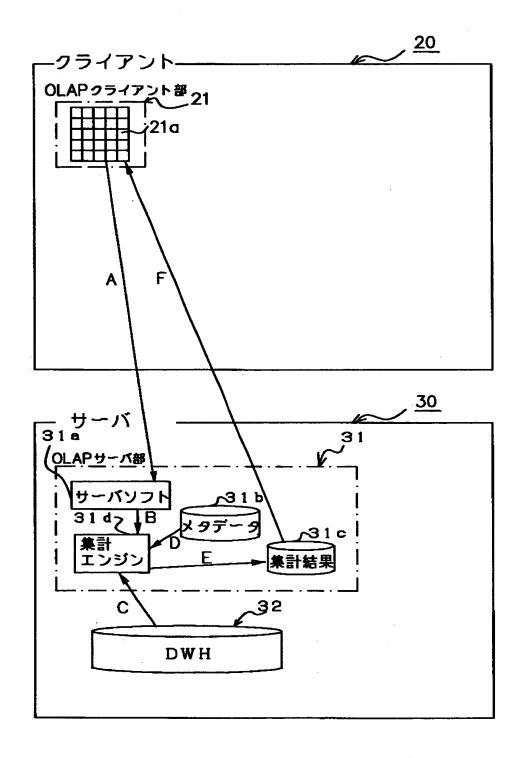
表頭項目(管理ポイント1)

表側項目(管理ポイント2)		X	Y	Z
	Α			
	В			
	U			
2)	D			

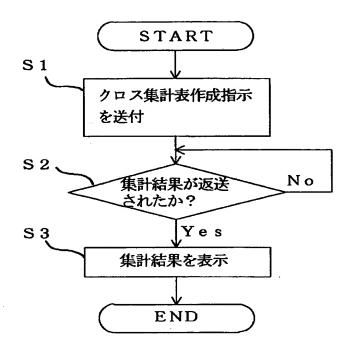
【図4】



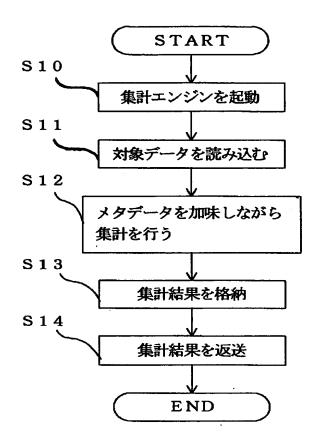
【図5】



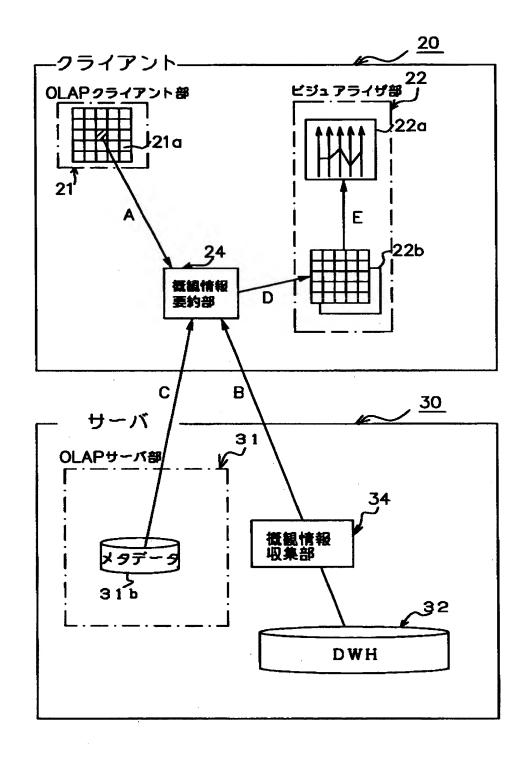
【図6】



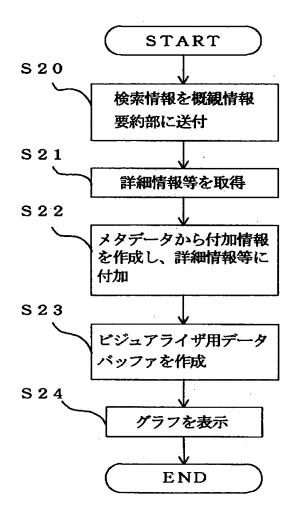
【図7】



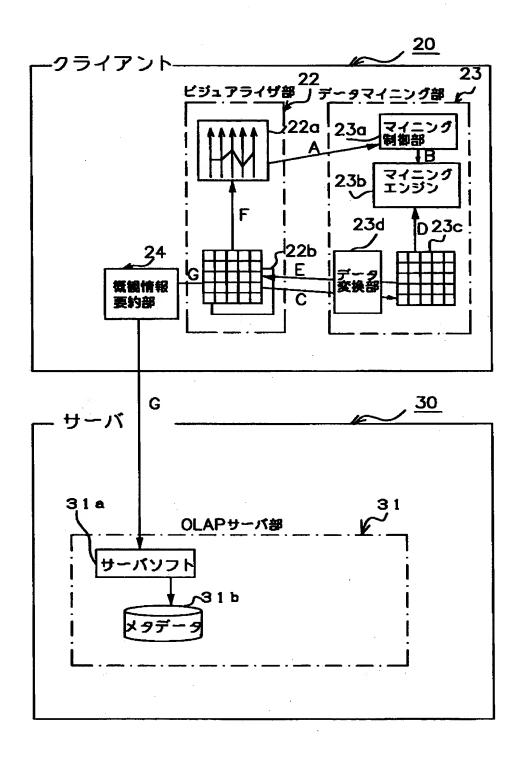
【図8】



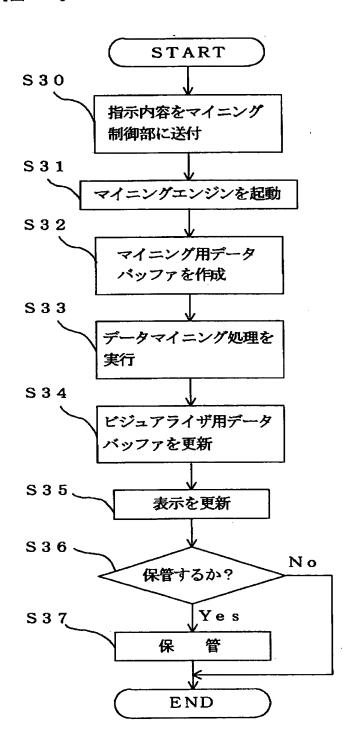
【図9】



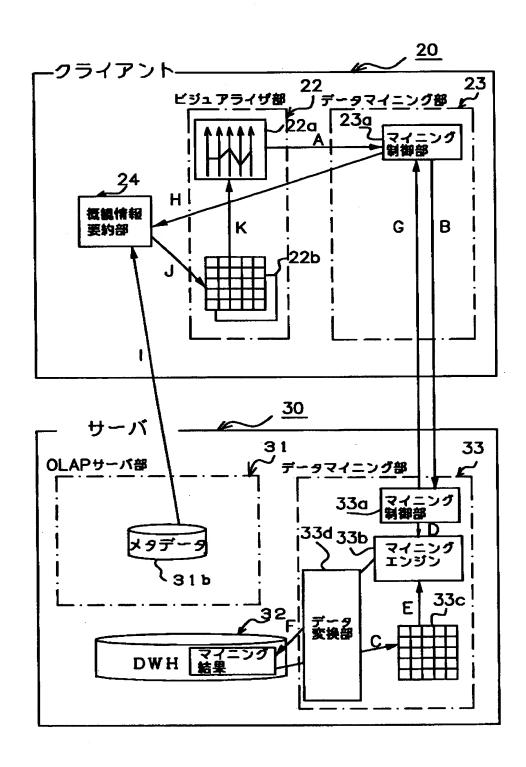
【図10】



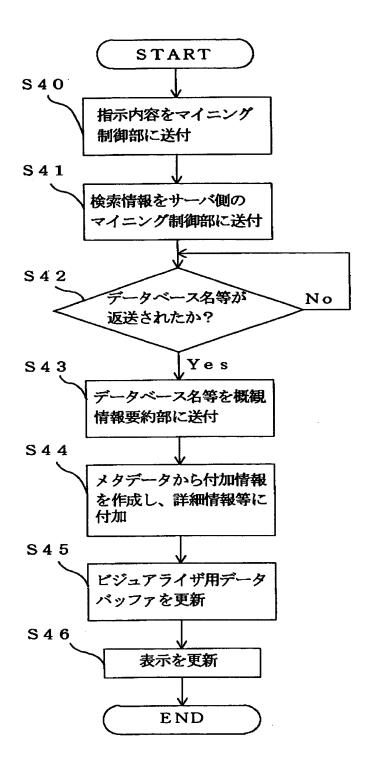
【図11】



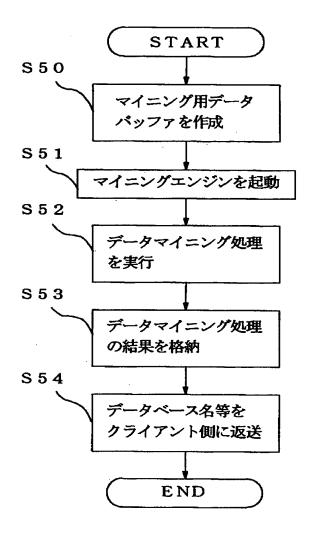
【図12】



【図13】



【図14】



【書類名】 要約書

【要約】

【課題】 エンドユーザがデータの分析を効率的かつ容易に行えるようにする。

【解決手段】 情報モデルがデータと独立したメタデータとして定義された方式で動作するOLAP21aをベースとして、集計されたデータの意味と原因探索の方向付けを、平行座標をベースにしたビジュアライザ22aによって行えるようにする。即ち、OLAP21aとビジュアライザ22aとが連係動作するようにし、分析者たるユーザが、OLAP21a或いはビジュアライザ22bの一方を操作すると、この操作が他方に反映されるようにすることで、直感的にデータの分析作業が行えるようにする。

【選択図】 図1

特平 9-163692

【書類名】

職権訂正データ

【訂正書類】

特許願

<認定情報・付加情報>

【特許出願人】

【識別番号】

000005223

【住所又は居所】

神奈川県川崎市中原区上小田中4丁目1番1号

【氏名又は名称】

富士通株式会社

【代理人】

申請人

【識別番号】

100072590

【住所又は居所】

神奈川県川崎市中原区上小田中4丁目1番1号 富

士通株式会社内

【氏名又は名称】

井桁 貞一

特平 9-163692

出願人履歷情報

識別番号

[000005223]

1. 変更年月日

1996年 3月26日

[変更理由]

住所変更

住 所

神奈川県川崎市中原区上小田中4丁目1番1号

氏 名

富士通株式会社